STUDIES IN INDUSTRIAL PSYCHOLOGY

BY

ELSIE OSCHRIN BREGMAN

NOV 3 1922

LC

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, in the Faculty of Philosophy, Columbia University

REPRINTED FROM

ARCHIVES OF PSYCHOLOGY

No. 59

STANFORD UNIVERSITY LIBRARIES

WITHDRAWN

3781
C72B

UNCLASSED
Dissertation

1922

STUDIES IN INDUSTRIAL PSYCHOLOGY

BY

ELSIE OSCHRIN BREGMAN

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, in the Faculty of Philosophy, Columbia University

REPRINTED FROM

ARCHIVES OF PSYCHOLOGY

No. 59

1922

The writer takes very gladly this opportunity to acknowledge her debt for much help and kindness to her teachers, Professors H. L. Hollingworth, R. S. Woodworth and E. L. Thorndike.

Professor Hollingworth, her first teacher, is largely responsible for her original interest in the subject of these studies, and his example, help, encouragement and very constant kindness have been a great factor in the progress of the work. To Mr. Percy S. Straus and Mr. A. S. Donaldson the writer is indebted for the opportunity to carry out the work herein reported, and to Professor A. T. Poffenberger for having kindly read and corrected part of the manuscript.

Finally, the many good people who were subjects in these experiments should not go unthanked for their patient performance of tasks that were often strange.

TABLE OF CONTENTS

Page

Foreword 4
Introduction 5

PART 1

Tests for General Ability 9

PART II

Chapter I
Tests for Special Abilities 25
Chapter II
The Negative Correlations 49
Chapter III
The Psychologist in Industry 55

APPENDIX

Scale 1 57
Scale 1A 59
Bibliography 59
Vita 60

FOREWORD

In April of the year 1919 the writer was employed by a department store in the city of New York to experiment in its organization with psychological tests. From the consequent investigations and use of tests these studies have been drawn.

At the time when this work was under way almost no work of a similar character had been undertaken. Very little appeared in the literature that bore directly on the problem, and organizations in which psychological tests, adequately evaluated and properly administered, functioned with any degree of importance, were hardly to be found.

Few of the many forms of tests which are now readily purchased in any quantity were then obtainable. Official and detailed information of the United States Army testing was not yet available. Other information on tests for and the testing of normal adults was very scant.

These circumstances, together with other conditions, more or less unavoidable, under which the studies were made, have operated to place certain limitations upon them.

The studies are not laboratory studies. They come out of the midst of a very busy industry. They were in consequence subject to restrictions, exercised not only by the factors of time and cost, but also certain other factors, inherent in the situation of the youngest of the sciences trying to work out its usefulness in an environment traditionally and actually so foreign to it.

The writer believes that notwithstanding their shortcomings the studies have a certain significance and usefulness, and in this belief they are presented to the reader in this volume.

Studies in Industrial Psychology

INTRODUCTION

In this country and at this time the term industrial psychology connotes chiefly the use of psychological tests for the prediction of ability, general or special, native or acquired, with a view to the most effective placement of individuals within an industrial organization upon the basis of such knowledge.

The title of this paper uses the term in this special sense, since the paper treats of tests for general and special abilities. The latter are specifically the abilities of salesclerk and clerical worker in a large department store.

This work was begun early in the year 1919. At that time the literature was almost barren of information bearing on the use of selective tests in industry. There are today, indeed, no great host of such studies, but there is gradually accumulating a body of information on the selection of workers by mental tests. New tests, more particularly those that measure personality traits rather than general intelligence, are being constructed and the technique of evaluating the tests is undergoing an interesting development.

In 1913 Münsterberg (1) published his Psychology and Industrial Efficiency, in which appeared the account of his experiments with motormen, marine officers and telephone operators.

Three years later, in 1916, appeared Professor H. L. Hollingworth's (2) Vocational Psychology, which treats of the problems and methods involved in vocational guidance and which summarizes the work up to that time by Lough, Lahy and McComas, with telephone operators, commercial students and typists.

The Journal of Applied Psychology was begun in 1917. In that year there appeared in this Journal articles by Burtt

(1) Münsterberg, Hugo, Psychology & Industrial Efficiency, 1913.
(2) Hollingworth, H. L., Vocational Psychology, 1916.

(3), Rogers (4) and Scott (5), each of which was concerned with the evaluation of tests for vocational guidance or selection.

In the following year Flanders (6) reports an investigation in an express company and Oschrin (7) in a department store. In the same year Link (8) records an investigation in a munitions factory. All three articles treat of the correlation of tests with ability as workers. Flanders found no significant correlation—Link and Oschrin found positive correlations between certain tests and their criteria.

During the year 1919 there were published two articles on trade tests, one by Toops and Chapman (9), the other by E. S. Robinson (10); an investigation by Henmon (11) on tests for aptitude for flying, and reports by Thurstone (12) on tests for telegraphers and for office clerks. Link's (13) Employment Psychology was published in the same year.

Burtt (14) in 1920 published an account of a very thorough investigation carried out upon operatives and clerical workers in a Canadian rubber factory. Marcus (15) in that year found a team of association tests to be superior in predictive value and administrative detail to a Civil Service examination. His subjects are Hollerith punch operatives.

(3) Burtt, H. E., Professor Munsterberg's Vocational Tests. J. Appl. Psych. 1917, I, 201-213.

(4) Rogers, H. W., Psychological Tests for Stenographers and Typewriters, J. Appl. Psych., 1917, I, 268-274.

(5) Scott, W. D., A Fourth Method in Checking Results in Vocational Selection, J. Appl. Psych., 1917, I, 61-66.

(6) Flanders, K. J., Mental Tests of a Group of Employed Men showing Correlation with Estimates Furnished by Employer, J. Appl. Psyc., 1918, II, 197-206.

(7) Oschrin, E., Vocational Tests for Retail Saleswomen, J. Appl. Psych. 1918, II, 148-155.

(8) Link, H. C., An Experiment in Employment Psychology, Psych. Review, 1918, XXV, 116-127.

(9) Toops, H. A. and Chapman, J. C., A Written Trade Test. Multiple Choice Method, J. Appl. Psych., 1919, III, 358-365.

(10) Robinson, E. S., The Analysis of Trade Ability, J. Appl. Psych., 1919, III, 352-357.

(11) Henmon, V. A. C., Air Service Tests of Aptitude for Flying, J. Appl. Psych., 1919, III, 103-109.

(12) Thurstone, L. L., Mental Test for Prospective Telegraphers, J. Appl. Psych., 1919, III, 110-117. A Standardized Test for Office Clerks, J. Appl. Psych., 1919, III, 248-251.

(13) Link, H. C., Employment Psychology, 1919.

(14) Burtt, H. E., Employment Psychology in the Rubber Industry, J. Appl. Psych., 1920, IV, 1-17.

(15) Marcus, L., Vocational Selection for Specialized Tasks, J. Appl. Psych., 1920, IV, 186-201.

Otis (16) examined mill workers and failed to get significant correlation between ability and the tests which he used.

In 1921 Freyd (17) published test correlations for journalistic aptitude. Bills (18) (19) tests for the selection of comptometer operators and stenographers, and Moore (20) a monograph on the personnel selection of graduate engineers. He constructed a test which differentiated between sales and design engineers, and found also that occupational interests show a definite correlation with the kind of occupation in which a man is successful. Bregman (21) first reported the correlations for salesclerks and clerical workers in a department store which are presented again in this paper.

Early in 1922 Ream (22) found a series of Downey will-temperament tests of positive value in predicting success in selling insurance.

Such is, in brief, a record of the investigations that have been made up to the present on the use of tests for selective purposes in industry. To what extent and in what manner they actually function in industry may be gathered from three reports published in consecutive years by the (23) National Association of Corporation Training. In 1921 thirty-nine organizations out of one hundred and seventy-two interrogated were using standardized tests in connection with employment, training and other personnel functions.

The data for the studies of this paper were collected in one such organization during the two years which the writer

(16) Otis, A. S., The Selection of Mill Workers by Mental Tests, J. Appl. Psych., 1920, IV, 339-341.

(17) Freyd, Max, A Test Series for Journalistic Aptitude, J. Appl. Psych., 1921, V, 46-57.

(18) Bills, M. A., Methods for the Selection of Comptometer Operators and Stenographers, J. Appl. Psych., 1921, V, 275-283.

(19) Bills, M. A., A Test for Use in the Selection of Stenographers, J. Appl. Psych., 1921, V, 373-377.

(20) Moore, B. V., Personnel Selection of Graduate Engineers, Psych. Monograph, 1921, XXX, Whole No. 132.

(21) Bregman, E. O., A Study in Industrial Psychology—Tests for Special Abilities, J. Appl. Psych., 1921, V, 127-151.

(22) Ream, M. J., Group Will Temperament Tests, J. Appl. Psych., 1922, XIII, 7-16.

(23) National Association of Corporation Schools. (Training).
1919 Psychological Tests and the Results Obtained therefrom—Seventh Annual Proceedings.
1920 Psychological Tests and Rating Scales in Industry, Eighth Annual Proceedings.
1921 The Application of Psychological Tests and Rating Scales in Industry, Ninth Annual Proceedings.

spent there as psychologist investigating the possibilities of selective tests.

PART I is a series of studies with two sets of Trabue's completion sentences. These were used mainly in testing applicants for employment for the purpose of identifying individuals, who, in general intelligence or ability, deviated from the normal for better or worse. A large number of people already employed were also tested with one of the two sets of sentences.

The Trabue sentences were chosen for this purpose of measuring general intelligence, since they correlate as closely, at least, as any other single test, with the longer and more elaborate tests which are used for this purpose, and because the use of any such lengthy and involved test was, under the circumstances, wholly out of the question.

From the testing of applicants for the period of about two years there accumulated a body of data, that together with the data from the tests of employees, has been analyzed and presented in PART I.

PART II is a history of the investigation the purpose of which was the establishment of tests that would indicate, from among outwardly undifferentiated applicants for employment those individuals who would do their best work as salesclerks, and those who should be employed for clerical work.

Chapter I of PART II appeared in the June 1921, number of the Journal of Applied Psychology under the title "A Study in Industrial Psychology—Tests for Special Abilities." A few unimportant changes have been made in the present copy.

PART I

TESTS FOR GENERAL ABILITY

Two Language Completion Scales

The tests which are the subject of these studies are 2 scales derived from Trabue's language completion exercises.* They were devised to test the applicants in the employment office of a large department store. Adults already employed were also tested with these scales for one purpose or another. The two tests will be called 1 and 1A throughout the paper. Sentences will be designated by their position in the test and the test number. Thus, sentence 1 of Scale 1A will be called 1A-1.

Derivation

Each test is made up of 25 of the Trabue sentences. The tests were so arranged as to approximate tasks equal in difficulty, using the difficulty values which are given in the original monograph. The order of sentences in each test is that of increasing difficulty, according to the following scheme.

1	sentence	with	a	scale	value	of	1
2	sentences	with	an	average	scale	value	of 2
2	"	"	"	"	"	"	" 3
2	"	"	"	"	"	"	" 4
2	"	"	"	"	"	"	" 5
3	"	"	"	"	"	"	" 6
3	"	"	"	"	"	"	" 7
3	"	"	"	"	"	"	" 8
3	"	"	"	"	"	"	" 9
2	"	"	"	"	"	"	" 10
2	"	"	"	"	"	"	" 11

More of the difficult sentences and less of the easier were used because the tests were to be used with adults whom it was unnecessary to test extensively in the lower values.

Subjects

Over a period of 21 months, 11,018 applicants were tested with the Scale 1A, and 4,053 were tested with Scale 1. This

*See Completion-Test Language Scales. Marion Rex Trabue.

number represents practically all of the applicants for positions during these months, with certain exceptions. The exceptions were mainly employed for such jobs as porters, dishwashers, scrubbers, milliners and tailors; i.e., unskilled and skilled manual workers, and the foreign born who were unable to read and speak English. Those who were tested were considered for such jobs as salesclerk, clerical worker, packer, cashier, wrapper, messenger, floorwalker, stockman, driver and driver's helper, and certain subordinate executive positions.

The following table shows the grade in school at time of leaving for the applicants who presented themselves during the first two and a half weeks of testing. This group may fairly be taken to be typical of all subsequent applicants. The range is from the 4th grade in the primary school into the professional schools and colleges. The mode is at the 8th grade.

Schooling Distribution—Table 1

Grade	4	5	6	7	8	1	2	3	4	Colleges, etc.
No.	2	1	5	34	75	21	10	6	20	6

With the ages of the applicants arranged in five year groups, the modal age group is from 15 to 19 years.

Both men and women were tested, but women were far in the majority. The applicants came from among the working class, the untrained working class of New York and its environs, and to some degree they were floaters of that class—women who were married and worked part time, girls who hoped to be married, unsuccessful storekeepers, men who had learned no trade, newly arrived immigrants from the English speaking countries. The second generation of most European races and countries were represented and a number of the first generation who had been here long enough to learn the language. Jews, Irish and Italians formed the greatest part of the foreign born or the foreign parented. There were some South Americans but no negroes.

The tests were given with a time limit of 10 minutes. This was decided upon after preliminary experimentation indicated that most subjects would attain their maximum performance in that time and that to very few would it be possible to make a perfect score in that period. No emphasis was placed upon speed in the directions for taking the test, which were always

oral and were followed by a short exercise similar to the test itself, namely, completing the two sentences, "One and —— are two," and "The boy lost —— hat and ——." The applicants were told to "——— finish as many sentences as you can, but do them carefully and correctly."

The customary precautions as to similarity of directions, accuracy of timing and constancy of conditions were, of course, observed. Applicants were tested either singly or in groups that did not exceed 10, according to the number in which they presented themselves for employment. The sentences were printed on the inside of a 4 leaf folder in large clear type. The sample sentences were on the outside sheet of this folder. The test in each case was begun only after the fore exercise had been correctly performed and it was evident that the task was clearly comprehended.

The sentences were scored in accordance with Trabue's scheme of 2 credits for each sentence perfectly completed, 1 credit for each almost perfect, and zero for any other performance. The maximum score was, therefore, 50. Trabue's Key for Completion Test Language Scales was used as a guide for marking.

Norms

Norms are available for 11,018 cases tested with Scale 1A and 4,053 tested with Scale 1. The two tests proved to be practically equivalent tasks, as is evidenced by Table 2, which gives the scores attained by 25%, 50% and 75% of each group, and Table 3 which shows the distribution by deciles.

TABLE 2*

Quartile Distribution

Test	Q1	M	Q3	No. of cases
1A	30	36	41	11,018
1	30	34	39	4,053

TABLE 3

Decile Distribution

	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
1A	26	29	32	34	36	38	40	42	44	50
1	25	29	31	32	34	36	38	40	42	50

No. of cases

1A 11,013

1 4,053

*In all Tables the whole number only of the step in which a measure falls is given. In computations, however, the exact fractional number has always been retained.

It is apparent that for all practical purposes the two tests are equivalent, although Scale 1 seems to be slightly more difficult.

Norms for Foreign Born

The performance norms for people of foreign birth differ from the norms of the native born. The figures in Table 4 were determined from 111 cases of foreign born who were able to read, speak and write English to such an extent that they were at least capable of understanding and following the test directions and were able to perform the test. These people had been in the United States for varying lengths of time, but all of them naturally and preferably spoke and thought in a language other than English.

TABLE 4—Foreign Born, Scale 1A

Quartile Distribution

Q1	M	Q3	*No of cases*
20	24	27	111

Decile Distribution

10%	20%	30%	40%	50%	60%	70%	80%	90%	100%	No. of cases
14	18	20	22	24	26	26	28	31	40	111

These scores are considerably lower than the scores of Tables 2 and 3, which give norms for the native born.

Five Minute Norms

Scale 1A was used as a five minute test with 440 individuals. Of this number 223 were applicants, 49 were salesclerks, 76 minor executives, department heads, and 30 were students at the N. Y. U. Training School for Teachers of Retail Selling. The median, quartile and decile scores for this group are given in Table 5.

TABLE 5

Score in 1A — 5 Minute Test

Q1	M	Q3	No. of cases
24	29	35	440

10%	20%	30%	40%	50%	60%	70%	80%	90%	100%	No. of cases
22	24	26	27	29	30	33	36	39	50	440

The scores are naturally lower than when the test is performed for ten minutes. In the range of the middle fifty per cent there is a constant difference of 6 points between the Q_1's, the M's and the Q_3's. Or, put in a different way—75% of the distribution of the ten minute test exceed the Median of the five minute test. There is also a greater interval, in the five minute test, between the Q_3 and the limit of the distribu-

tion, making possible a finer discrimination of individuals above 75% in the shorter time period.

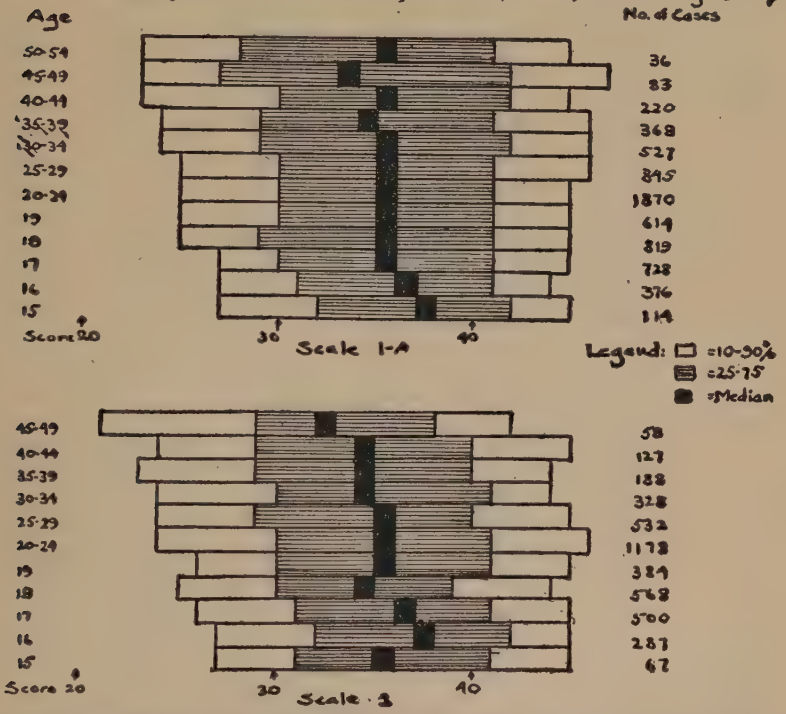
Age Norms

Age distributions for both ten minute tests were plotted in one year periods from 15 to 20 years, and five year periods from then on. The upper age limit is 55.

Figure 1 presents the data graphically and Tables 6 and 7 numerically.

Figure 1

Showing Scores Attained by 10, 25, 50, 75 + 90% of Each Age Group



The medium and quartile measures show no progressive appreciation or depreciation with age. Neither does the point indicating the uppermost decile vary with age but remains practically stable throughout. The point indicating the lowest decile, however, falls off slightly but steadily as age increases. Just what this means it is difficult to say. A possible explanation is that there is a tendency toward a slight increase, with age, in the number of low grade people looking for employment. Another possibility may be that there is an actual

TABLE 6

Age Norms for Test 1A—10 Minute.

Age	10%	25%	50%	75%	90%	No.
15	27	32	37	41	44	114
16	27	31	36	40	43	376
17	25	30	35	40	44	728
18	25	29	35	40	44	819
19	25	30	35	40	44	614
20-24	25	30	35	40	44	1870
25-29	24	30	35	40	45	845
30-34	24	29	35	41	45	527
35-39	24	29	34	40	45	368
40-44	23	30	35	41	44	220
45-49	23	27	33	41	46	83
50-54	23	28	35	40	44	36

TABLE 7

Age Norms for Test 1—10 Minute.

Age	10%	25%	50%	75%	90%	No.
15	27	31	35	40	44	67
16	27	32	37	41	44	287
17	26	31	36	40	44	500
18	25	30	34	38	43	568
19	26	30	35	40	44	384
20-24	24	30	35	40	45	1178
25-29	24	29	35	39	44	532
30-34	24	30	34	40	43	328
35-39	23	29	34	39	43	188
40-44	24	29	34	39	44	127
45-49	21	29	32	37	41	58
50-54			34			16

waning of ability with age, but that it is confined only to those individuals who fall at the lowest extreme of the curve of distribution.

Group Differences

Scale 1A was used as a five minute test with various special groups among the personnel of the store and for a short period also for applicants. A class at N. Y. U. Training School for Teachers of Retail Selling was also so tested. Interesting dif-

ferences are to be noted between the performances of these groups. See Table 8.

TABLE 8

Performance of Groups Tested with Scale 1A—5 Minute.

	Q1	M	Q3	No.
Applicants	24	28	31	223
Sales clerks	24	29	33	49
Sec. Mgrs.	24	29	34	63
Executives	26	34	38	62
Students	34	39	44	30

The performance of the applicants, sales clerks and section managers—or "floorwalkers", as they are generally known—is very much alike, except that above the median the two latter groups are very slightly better. There is a marked difference between these groups and the executives, however. Nearly 70% of the executives exceed in score the median of the sales-clerks and the section managers. The students, in turn, make better scores than the executives, for, of the students, 75% reach or exceed the median score of the executives.

Differences Within a Group

The executives noted in the preceding section occupied leading positions in the organization and as such represented roughly the most able individuals of a group of 5,000, from the point of view of business success.

Ratings were made on 51 of these executives by the general manager and by one of the assistant managers, who, by virtue of his function in the organization, was in a position to know the department heads unusually well.

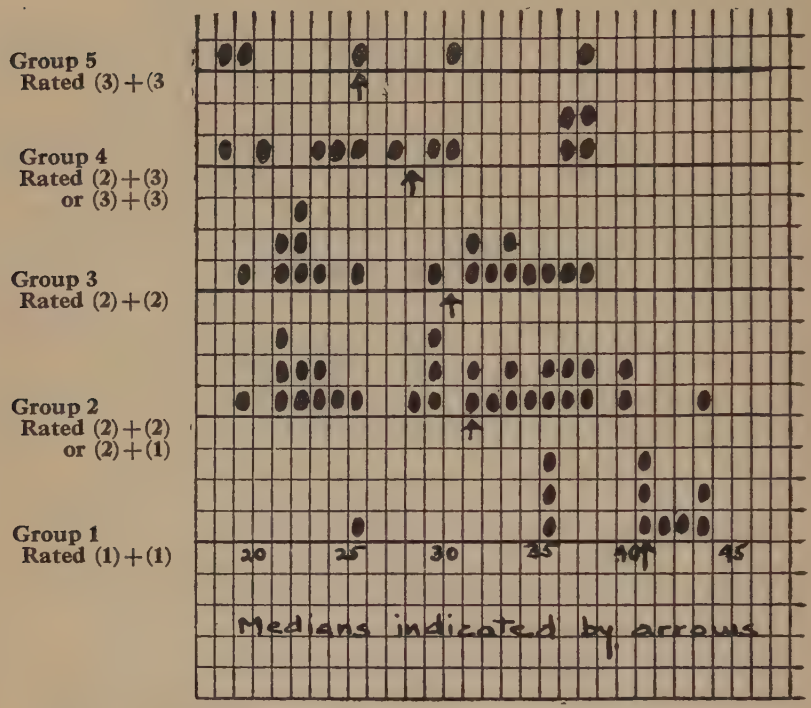
These two men classified the executives in three groups, (1) the most able, (3) the least able, and (2) those of average ability from the standpoint not only of their present position and accomplishments, but also of their possibilities for further development. The ratings of the two men were combined in the following fashion:

Group 1 Those classified as (1) by both raters
Group 2 Those rated either as (2) (2) or (2) (1)
Group 3 Those rated as (2) by both raters
Group 4 Those rated either as (2) (3) or (3) (3)
Group 5 Those rated as (3) by both raters.

Groups 1, 3 and 5, it is apparent, possess greater validity than Groups 2 and 4.

The scores for these groups are shown in Figure 2.

Figure 2—Executive Scores in Scale 1A—5 Minute, and Ratings



Although there is much overlapping, the differences in the performance of these groups may be noted. Group 1, the first rate executives, make a median score of 41, fully 9 points above the median of Groups 2 and 3, who are the executives of average ability. These, in turn, score 5 points higher than Group 5, which both judges agree upon rating as inferior.

From the data of this and proceding sections, it would seem that some of the same factors are involved in the performance of a completion language test as operated to keep some people salesclerks or floorwalkers, make others executives, and of these, executives of varying ability, and lead still others to continue their studies beyond the college or professional school.

These factors, it will probably be generally agreed, constitute what has been summed up under the term "general intelligence".

Correlation with Otis Advanced Examination

A number of individuals who had been tested with Scale 1 or 1A were also tested with the Otis Advanced Examination. Correlation coefficients between the Otis and the two completion scales are shown in Table 9. The median, average and S. D. from the average for each variable are also given in this Table for the purposes of comparison. Comparison is possible both between the groups whose members were tested with both the Otis and a completion test and between these groups and the total population tested.

Groups 1 and 2 of the Table were candidates for a training class for executives. Group 1, numbering 40, were tested with Scale 1A for 10 minutes; Group 2, numbering 13, performed this test for 5 minutes only. Both groups are selected on the basis of ability, for candidates for this training course are either already in positions of some responsibility or, as clerks, have shown latent ability which the course is to develop.

Comparison of the medians for these two groups in the 1A test with the medians given for this test for the total population tested—to be found in Tables 2 and 5—show that in test performance also this is a superior group.

Group 3, which performed the Scale 1—10 minutes, were messengers and wrappers and the like, most of them between 15 and 21 years of age. This group also includes some individuals who were tested because of reported inefficiency or suspected mental defect. The difference between the average of this group and that of Groups 1 and 2 in the Otis test is very great. In Scale 1, however, there is a falling off of only one point from the median of the total unselected group, (cf. Table 2). Group 3 then is apparently much inferior to Groups 1 and 2, but approximates in test score the average of the total population tested.

Group 4 is composed of 103 school children in the 8th grade of a New York City public school. They form, of course, a group highly selected, both as to age and ability. In test performance their position is between Groups 1 and 2 and Group 3.

TABLE 9

*Correlation of *Otis Advanced Examination with Scales 1 and 1A*

Group	r	Variables		Median		Mean		S. D.		No.
		x	y	y	x	y	x	y	
1.	.804	Otis	1A-10′	40	136.3	39.3	38.6	7.4	40
2.	.71	Otis	1A-5′	34	130	33.9	34.8	6.2	15
3.	.846	Otis	1 -10′	33	90.8	32.8	35.3	8.3	40
4.	.49	Otis	1A-10′	34	106.4	34.1	19.9	4.4	103

The correlation throughout is high and positive. The coefficients for the adult groups are very high; r = .804, .71 and .846 respectively. Moreover, the coefficients for groups 1 and 3 are very close. That the correlation for group 2 differs as much as it does from these two is probably due in part to the small number of cases in this group and the consequent larger unreliability, and also to the fact that as a five minute test, Scale 1A does not give so true a measure of ability as it does as a ten minute test.

The smallest coefficient comes from the group of school children. But the variability of this group is about one-half what it is for the other groups. (Using Thorndike's formula $\frac{\text{Var}}{\sqrt{\text{C. T.}}}$ † Group 4 in the Otis test is almost exactly one-half as variable as Group 3. The exact ratio is $\frac{1.9}{3.7}$.) Since, when the variability of group relative to the C. T. is smaller, the correlation coefficient is also smaller, the reduced size of the coefficient for Group 4 can be thus accounted for.

Difficulty Value of Sentences

In arranging the two tests, the sentences were placed in an order of increasing difficulty, according to the values determined by Trabue from the groups which he tested. These groups were made up of children from the second grade through High School and a small number of graduate studdents. It seemed reasonable to suppose that a group of adults only would give values somewhat different from those drawn almost wholly from children, both as to the absolute difficulty of the sentences and the order of their difficulty.

In order to determine this the difficulty of each sentence used in the two series was calculated in terms of the P. E.

*In some cases the Otis scale was abbreviated by omitting tests 5 or 5 and 6. Full scores were computed for such cases by means of the formula and values given by Otis for this purpose on p. 14 of the Manual of Directions, 1921 revision of the Otis Group Intelligence Scale.

†E. L. Thorndike. Mental and Social Measurements, p. 133.

from the median—using the first thousand cases tested with each series of sentences.

Trabue's values are also based upon the P. E. and the two sets of values are, therefore, comparable. An important distinction between the two sets must, however, not fail to be pointed out. Trabue's subjects were, with the exception of the small group of graduate students, all children in the public schools from the 2nd grade through the High School. The performance of these subjects varied from grade to grade, each higher grade showing an increment in score. There was, therefore, a separate distribution for each grade—and Trabue's final difficulty value for any sentence is an average of the P. E. positions of that sentence in the grades in which it was used. Into the minutae of the derivation of the values it is not necessary to go here; suffice it to point out that Trabue's values are not simple P. E. positions but composite figures based on several such positions.

In the work here reported, however, there was only one distribution to be considered; the performance of adults varies very little with age as has been shown in a preceding section. The values here determined are, therefore, true P. E. positions, whereas the original set of values are, as their author has called them, general locations determined from several P. E. positions in several distributions.

The original values, moreover, are expressed, not in units above and below the median but in units above an arbitrary zero point. In order to facilitate comparison of the two sets of values, the new values have been referred to the same zero point by giving to the easiest sentence of the fifty used in the two scales the same value as appears for it in Trabue's monograph. This sentence is 1A-1, its position by our data is 4.083 P. E. below the median. Trabue's value for this sentence is 1.38 and it is assigned the same value. All other values are computed from this as a base by adding to 1.38 the number of P. E.'s which the other sentences are distant from 1A-1 by our data.

The two series have by this device a common zero point, that established by Trabue. This probably more closely approaches the true zero than any point that could be determined from the data of this paper since Trabue's subjects included very young children, and these may more reasonably

be expected to approximate the point of absolutely no ability to perform a test of this character than adults.

In Table 10 are presented the values so obtained. The sentences are listed from the easiest to the most difficult as determined by the adult performances. The value of the sentences for these individuals in terms of P. E. from the median is shown, the computed difficulty of each sentence when 1A-1 is assigned the value of 1.38, which is identical with the Trabue value for this sentence and finally the Trabue value and number for each sentence.

It will be seen from this arrangement of the data that the order of difficulty of the sentences is not greatly changed. The correlation (ρ) between the two rank orders is +.97 when calculated by the Spearman formula $\rho = 1 - \frac{6\Sigma D^2}{n(n^2-1)}$ The range, however, is decidedly wider with Trabue's values, or it may be that the units are smaller. Sentence 25, scale 1A, which is the most difficult in both determinations, is 12.65 units above zero in Trabue's values, only 8.57 units in the adult distribution. Since both series begin with 1.38 the sentences with the adult group have a range of only 7 units, whereas with the earlier work the range is about 11 units, more than half as much again.

TABLE 10

Difficulties of Sentences of Scales 1 and 1A

Sent. No.		P. E. from Median	Computed Difficulty Value	Trabue's Value	Trabue's Sent. No.
1A	1	–4.083	1.38	1.38	1
1	1	–3.506	1.96	.96	2
1A	5	–3.450	2.01	4.47	12
1A	2	–3.3	2.15	1.63	5
1	2	–3.015	2.45	2.52	75
1A	3	–2.986	2.48	1.97	77
1	4	–2.986	2.48	3.34	11
1	3	–2.932	2.53	1.09	76
1A	6	–2.905	2.56	3.66	29
1A	4	–2.834	2.63	3.31	7
1A	7	–2.631	2.83	4.15	22
1	5	–2.514	2.95	3.58	19
1	7	–2.514	2.95	4.26	16

TABLE 10 Continued

Sent. No.	P. E. from Median	Computed Difficulty Value	Trabue's Value	Trabue's Sent. No.
1 9	–2.425	3.04	5.55	30
1 11	–2.384	3.08	5.85	98
1A 12	–2.211	3.25	6.15	31
1A 8	–2.166	3.30	5.69	23
1A 11	–2.074	3.39	6.16	61
1 8	-2.054	3.41	5.40	24
1 6	–1.835	3.63	4.12	17
1A 9	–1.780	3.68	6.95	58
1A 13	–1.780	3.68	7.31	27
1A 10	–1.512	3.95	5.98	25
1A 14	–1.462	4.00	7.85	107
1 14	–1.374	4.09	6.96	37
1 15	–1.368	4.10	7.04	69
1 10	–1.286	4.18	6.32	57
1A 15	–1.253	4.21	7.16	28
1 16	–1.110	4.35	8.32	94
1 13	–1.000	4.46	6.67	34
1A 17	– .740	4.72	7.91	41
1 12	– .592	4.87	6.50	102
1A 16	– .504	4.96	8.15	36
1 17	– .496	4.97	8.37	70
1A 18	– .330	5.13	8.29	68
1 18	– .130	5.33	8.28	43
1 19	.307	5.77	8.91	62
1A 19	.357	5.82	8.92	32
1A 20	.480	5.94	8.92	45
1 20	.489	5.95	9.04	81
1A 21	.644	6.11	9.28	53
1A 22	.958	6.42	9.53	91
1 22	1.279	6.74	9.88	93
1A 23	1.318	6.78	10.14	55
1 21	1.884	7.35	10.05	50
1 23	2.044	7.51	10.19	78
1A 24	2.103	7.57	11.58	83
1 25	2.155	7.62	11.14	86
1 24	2.357	7.82	10.55	85
1A 25	3.111	8.57	12.65	88

Summary

Two language completion tests, called Scale 1 and 1A respectively, each composed of 25 Trabue sentences, are the subject of these studies. The data concerning the tests are derived from the performance of adults, applicants for employment and employees of a large department store.

I. Norms are given for each test when used with a ten minute time limit, which are derived from the performance of a total of above 15,000 adults for both tests.

II. The two tests appear from these norms to be of practically equal difficulty.

III. Norms for 111 individuals of foreign birth, who performed Scale 1A are given. These norms, which are considerably lower than the performance of the native born, illustrate the fact that to some extent proficiency in the tests is dependent upon familiarity with the English language.

IV. Norms for 440 individuals who performed Scale 1A for a five minute period only are also given.

V. Age distributions covering a period of years from 15 to 55 for Scale 1A and from 15 to 50 for Scale 1 are given. No constant change in performance with age is to be observed except at the lower end of the curve. The score tends here to decrease with age.

VI. The difficulty of each sentence in terms of probable error of the distribution, using 1,000 cases as data, is given. Compared with the difficulty derived by Trabue for the same sentences, the order of difficulty is found to be little changed—ρ being .97. The absolute values are different, however. The sentences, with Trabue's subjects and units, cover about 11 units, with the adults the sentences cover only 7 units.

VII. Differences in the performances of several occupational groups are found that conform generally to the rank of these groups.

VIII. Differences in the performance of executives are also found that follow the ability of these executives according to managers' ratings.

IX. Both scales give high correlation with the Otis advanced examination.

$r = .804$ for Otis and Scale 1A–10′
$r = .846$ for Otis and Scale 1 –10′
$r = .71$ for Otis and Scale 1A–5′

The correlation for Scale 1A-10 minute with Otis when children of the 8th grade of a New York City public school are used as subjects is not so high; r = .49. The greater homogeneity of the group of schoolchildren is considered responsible for the reduction.

Conclusion

That in the performance of both tests information bearing on the general intelligence, or what is perhaps a better term, the "general ability" of the person tested, is given, seems to be indicated by the close correlation with the Otis examination, and by the positive association with executive ability and general success in life as measured by occupational rank.

For this purpose of gaining information concerning the intellectual quality of the individual tested, the two tests, Scale 1 and 1A, were found to be very serviceable in the organization in which they were used. By selecting against, from among the individuals, those applicants who made scores toward the lower end of the curve of distribution, employment of the dull and defective could be guarded against.

The critical score, below which employment was questionable, could be varied with the needs of the job and had to be varied with the urgency of the employment situation. The usefulness of the test was not limited, however, to the detection of the individuals below the average. Information concerning the above average individuals was also necessary and useful. And the tests served these purposes not only for applicants, but for employees also, in cases of unsatisfactory work or behavior or of promotion, transfer, or selection for training in the classes of the educational department.

The tests do not, of course, furnish very fine measures of mental age or ability. But such measures are not generally needed for the bulk of the workers. To have information that roughly classifies a person as average, above or below average, or very much above or below average, is in most cases valuable and sufficient information. Further detail would concern itself with that person's special abilities rather than general ability.

For cases where a more refined measure is needed, longer tests are generally available for a more accurate rating. Even in such cases the short test, given first, has served as a guide as to which of the longer tests to use.

Aside from the fact that they are indicators of general intellectual level, certain qualities appertaining to the tests make them especially fitted for industrial use. Not the least of these is that the test period is a short one. For use in an employment office it is especially desirable that this should be so; on the one hand in the interest of the management for whom it is desirable from the point of view of costs and administration that the business of employment be as simple and brief as possible; on the other hand, in the interest of the applicant, for whom the total time spent without remuneration, in an interview, a mental examination, and often, in addition, a physical examination takes up no mean part of a working day.

Besides being short the tests are easy to give. An intelligent clerk, properly trained, can give them. The tests are readily understood, even by the very dull, and the task seems not too strange and fantastic to people to whom the fact of any kind of a test is surprising.

The marking of the tests is also not difficult, and with practice becomes very rapid indeed, although the use of a stencil is, of course, not possible. That by reason of the nature of the tests a certain amount of personal judgment may influence the marking is a weakness, as is also the dependence of performance to a certain degree upon proficiency in the English language, which was illustrated by the lower scores for the foreign born. For the latter fact, due allowances can always be made if the country of birth is indicated; as for the former it was found in practice, and with practice, that whatever was the error that personal judgment introduced into the scoring, it was at least not large enough to result in a gross misjudgment of the intelligence on the basis of the score.

PART II

TESTS FOR SPECIAL ABILITIES

CHAPTER 1

The greatest number of employees of a department store are salesclerks; the next in number are clerical workers. Consequently it was for tests that would indicate these two types of workers that investigation was first begun.

It has seemed worth while, in the light of the many new problems of administration which confront the investigator in industry, to add to the report of the findings of the investigation, an account of the manner in which the investigation itself was handled, since this phase is as vital to the success of the undertaking as an adequate knowledge of the scientific technique and procedure.

Preliminaries. In April of 1919 the writer came to the department store, whose employees number 5,000, to "experiment with vocational tests." Before beginning experimental work of any kind the writer went through a course of training, working in every department of the organization in every capacity—as wrapper, salesclerk, complaint clerk, etc., etc., through a miscellany of positions. This course enabled the writer not only to understand the organization in which she was to work, but to learn the personnel and personality of departments, to have more than a bowing acquaintance with department heads and executives, and, not the least important, to have a knowledge of useful kinds of records and of many sources of information which blueprints and verbal reviews did not give.

Immediately following this "course of sprouts," the writer became an active member of the Employment Department, interviewed and appointed applicants and in three months (during which time an office which could be used as a laboratory was being built and equipped, and preliminaries to actual experimental testing, such as collection and printing of forms, etc., were carried on) acquired a grasp of the local labor situation, the sources of supply, the types, characteristics and social

levels of applicants who presented themselves, the ways in which appointments and judgments were made, etc., etc.

The following diagram illustrates the office laboratory which was built where the testing could be carried on with the minimum of noise and disturbance.

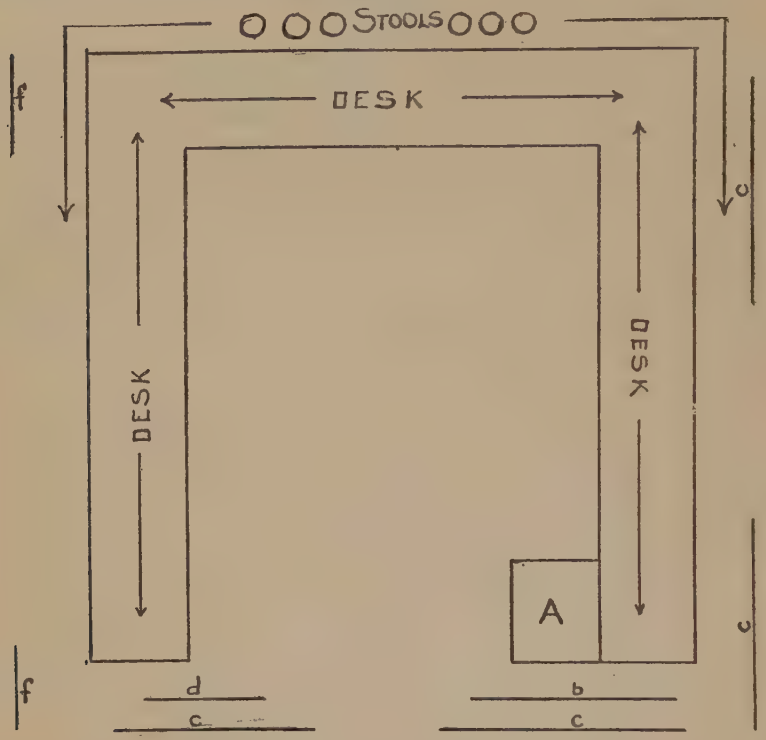


Fig. 3. Plan of office laboratory.

a. Examiner's Desk
b. Bookcase
c. Windows
d. Blackboard
e. Cupboard
f. Doors

Before the testing itself took place, meetings of the executives, heads and assistant heads were called, at which the purpose, plans and method of the investigation were disclosed and explained as simply as possible, and cooperation was urged. In order to dispel prevalent bugaboos about the nature of mental tests, these executives were tested with Scale 1A 5 minutes.

Obtaining the Criteria. In view of the unfortunate waste that has not infrequently occurred when after extensive tests had been made, records of work and ability were unobtainable,

it seemed wise to secure the criteria before testing, the more so that some coherent scheme of rating had to be worked out in an organization where for two thousand salesclerks there were as many as 200 department and section heads.

Salesclerks. Considering first the salesclerks, two methods of determining their ability were available (1) the actual weekly sales or production record made by each clerk, (2) the opinions of department heads, of which there are for each department

1. Buyer.
2. Assistant Buyer.
3. Section Manager or Floorman.

Production Record. Since the gross amount of sales varies for each department and from week to week, according to season and weather, it was not possible to throw all sales on to one scale for all time. The following method was used.

A simple frequency table was made showing the sales for each department for each week. From this the upper and lower quartile limits were determined, and record was made for each clerk, whether he sold within the interquartile range, above it, or below it. Such a record was kept for ten weeks and from this was determined whether a clerk sold characteristically within, above or below the average range.

Ratings. This information was supplemented by the ratings of department heads, which were obtained in the following manner:

The accompanying form was sent to each buyer and assistant buyer and section manager. From the information so gained it became possible to group the bulk of the selling force into three large categories—the good, the poor and the average salesclerks.

To M Sec. Man.
Buyer Dept.

Please send to ..... of the Employment Office, on this form, not later than ....., 19 ...., three lists of employees, carefully made out according to the following instructions.

I. List below the names of as many salesclerks as you know who are now employees of ——————————and who represent, in your opinion, the most capable and desirable type of salesclerk.

Write down *only* the names of *absolutely first class* salesclerks; those who really are able to sell.

No.	NAME	Remarks

II. List below the names of as many salesclerks as you know who are exactly the opposite of the people you have just named, representing the poorest type of salesclerk with little or no selling ability.

No.	NAME	Remarks

III. Place in blank (see opposite page) a list of salesclerks who are, in your opinion, neither very good nor very poor; the ordinary type of salesclerk who is satisfactory but not exceptional.

This loose form of rating sheet was used in preference to a more rigid and detailed form for the following reasons:

(1) The department heads of the organization under consideration are to a great degree free agents. It was essential for the return of the largest number of these sheets with a

No.	NAME	Remarks

minimum amount of energy and friction entailed, that the form be filled out with the utmost ease.

(2) The department heads are not an academic group—the median intelligence lower than that of such a group—they had consistently shown an aversion to any sort of analytic thinking or patient arrangement of groups into rank order, and such thinking, if forced upon them would be of a doubtful reliability. It seemed on the whole to be wisest to use a spontaneous method of rating—the more so that it would have been impracticable to test the entire force of salesclerks. Therefore, by this method of rating, the testing was confined to those individuals who were spontaneously recalled in the minds of the several department heads as meriting good, bad or indifferent ratings.

The results of this method bore out the choice of the course. The forms were returned on time and with a minimum of follow-up work. The large body of salesclerks was rated and the undertaking did not lose in prestige by being called super-scientific, impracticable etc.

The judgments so obtained from these three sources, buyer, assistant buyer and section manager, were then matched up on the form on page 30.

All the ratings were tabulated on this form. Against each salesclerk's name was placed in the appropriate column a red check, if called poor, a green if good and a black if average. The different rankings were further marked by placing red in the lower boxes only, green in the upper and black in the middle. Finally, under sales in the first column were placed

Department .....

Disc. No.	NAME	Sales	Buyer	Ass't Buyer	Section Manager	Remarks

tion record by which the judgments of department heads could be controlled, the clerks being engaged in a multiplicity of operations, from complicated statistical work and bookkeeping, to counting of checks, filing and comptometer operating. Moreover in most cases the clerks were well known to only one person, the local supervisor, only superficially known to the actual department head and very occasionally known to as many as three persons.

Raters were therefore instructed to rate only those clerks whom they knew and the rates taken at their face value, with the precaution that such clerks about whom there was a wide divergence of opinion were excluded from the possible subjects for experiment, and only such clerks included about whom at least two people were agreed.

The following scale indicates the manner in which the clerks were grouped.

1. Rated as first class by two or more—no dissenting opinion.
2. Rated as first by two—2nd by 1.
3. Rated as first by one—2nd by 1.
4. Rated as second by two—1st by 1.
5. Rated as second by two or three—no dissenting opinion.
6. Rated as third by one—2nd by 2.
7. Rated as three by one and 2nd by 1.
8. Rated as three by two and 2nd by 1.
9. Rated as three by three or two—no dissenting opinion.

In obtaining ratings throughout the course of this work, it was noted that there was a very marked reluctance to say that an employee was third rate or unsatisfactory, for the reason that the executive giving such a rating laid himself open to the question of why he retained such a person. It seemed fair, therefore, to infer that all clerks whose total rating was more than seven, could be included among the third rate clerks, and they were so considered. This was also done because the number nine group alone made up a very inconsiderable number. The ratings of *good* clerks, if unanimous, were significant; of the poor and mediocre clerks, less so because of the hesitancy to call any one poor, and of a tendency to underrate if the clerk was unknown or doing unimportant work, or if but newly employed.

However, these ratings, such as they were, served to indicate three groups, the very good (1) the poor (7-9) and the

generally average clerks who were rated between these two extremes.

Tests. Through the courtesy of Prof. E. L. Thorndike and the National Research Council Committee sets of the N. R. C. Tests, the precursors of the National Intelligence Tests were made available for the purposes of this investigation.

These tests were:

1. Verbal A which consisted of
   1. Arithmetic
   2. Directions
   3. Sentences
   4. Synonym Antonym
   5. Judgment
   6. Analogies
2. Non Verbal A
   1. Picture Completion.
   2. Series Completion
   3. Comparison
   4. Symbol Digit
   5. Form Combination
3. Verbal B
   1. Computation
   2. Vocabulary
   3. Sentence Completion
   4. Disarranged Sentences
   5. Logical Selection
4. Non Verbal B
   1. Copying Designs
   2. Pictorial Sequence
   3. Pictorial Identities
   4. Recognitive Memories

In addition the following tests were also used,

[1]Sentence Completion 10′ and 5′ time limits—called T1A

Arithmetic—Woody McCall Mixed Fundamentals—called T2

Rearrangement of Animals—An abbreviated form of the writer's test published in June, 1918, *Journal of Applied Psychology*—called T3

Woodworth Wells Mixed Relations

Woodworth Wells Mixed Relations changed to an underlining test—called T5

Woodworth Wells Opposites

Woodworth Wells Opposites changed to an underlining test—called T4

Woodworth Wells—Number group checking

Woodworth Wells—Cancellation

Testing. Directions for performing the tests were identical for every individual.

The tests were given to groups in the laboratory sketched above. The testing was always in the early morning—from

1. This is the test called Scale 1A in Part I.

9 to 10.30 a. m. Meetings did not last longer than 90 minutes. Subjects were recalled for a second period if the tests were not completed in the first.

Before beginning any testing, the purpose and method of the work was carefully explained to each group as simply as possible. Discussion was perfectly free and testing was never begun without good feeling and rapport between the subjects and the experimenter. Pains were taken to dispel any preliminary nervousness, and to see that all subjects took the tests comfortably. Both sexes and all ages from 17 to 60 were represented. Of the salesclerks there were sellers of every type of merchandise from upholstery and women's suits to shoes, veiling, gloves and notions. Every phase of clerical work at which the employees of this establishment were engaged was likewise represented in the test groups. There were skilled typists and stenographers, statistical workers, bookkeepers, billing clerks, comptometer operators, check counters, filers, keepers of simple records, etc., etc.

For the initial phase of the experiment, in which a rough evaluation of the tests as to their sensitivity to different degrees of abilities was sought, only the extremes of both groups were used; that is, of the salesclerks, those were tested who were *consistently* classed as good and as poor; of the clerical workers, those who, according to the classification scheme indicated above, were classified as 1, and those who fell within the 7 to 9 classifications.

The average workers of each group were made use of in a later phase of the experiment.

Correlation Formula and Statistical Methods. Several correlation methods were used in the course of this work.

1. The method of Unlike Signed Pairs—discussed in Thorndike's Mental and Social Measurements, p. 162, pp. 170-1—and called in tabulations and discussions in this paper the U Formula.

2. Pearson Biserial R.—a formula which determines the correlation when one variable is measured and continuous, the other unmeasured and alternative—and which closely approximates the $r = \frac{\Sigma xy}{n\sigma X \sigma y}$ see *Biometrika*, Vol. VII, 1909, p. 96.

3. Finally the orthodox Pearson $r = \frac{\Sigma xy}{n\sigma x\sigma y}$ was used in the last phase of the experiment in partial correlation and regression in order to weigh the significant tests. When this formula was used the value 1 was given to a quality i.e., good salesmanship, and 0 to the absence of it.[2]

The other formulas do not necessitate so arbitrary an assignment of values.

Correlation Coefficients—Salesclerks. Table II below represents the coefficients which were obtained from the salesclerks when the good and the poor clerks were tested with a series of tests. The number of clerks tested, the names of the tests

TABLE 11

GOOD AND POOR SALESCLERKS

Number of Cases	Tests	r by U Formula
45	N R C – N V A Total	— .09
46	" " " " " " 1 Picture Completion	— .39
44	" " " " " " 2 Series Completion	+ .22
44	" " " " " " 3 Comparison 3′	— .59
44	" " " " " " 3 Comparison 2′	— .83
43	" " " " " " 4 Symbol Digit	+ .25
44	" " " " " " 5 Form Combination	— .37
44	" " " V A Total	— .59
43	" " " " " 1 Math.	— .16
44	" " " " " 2 Directions	— .66
43	" " " " " 3 Sentences	— .79
44	" " " " " 4 Synonym Antonym	— .22
45	" " " " " 5 Judgment	— .79
43	" " " " " 6 Analogies	— .54
46	" " " N V B Total	+ .16
48	" " " " " " 1 Copying Designs	— .06
46	" " " " " " 2 Pictorial Sequence	+ .16
48	" " " " " " 3 Pictorial Identities	.00
45	" " " " " " 4 Recognitive Memories	+ .51
48	" " " V B Total	— .06
49	" " " " " 1 Computation	— .16
49	" " " " " 2 Vocabulary	.00
49	" " " " " 3 Sentence Completion	— .34
49	" " " " " 4 Disarranged Sentences	— .28
49	" " " " " 5 Logical Selection	.00
48	T 1 A Sentence Completion 10′	— .13
48	T 1 A Sentence Completion 5′	— .37
33	T 2 Arith.	— .16
49	W. W. Opposites	— .16
47	T 4 W. W. Opp. Adapted	— .43
48	W. W. Mixed Relations	— .37
48	T 5 W. W. M. R. Adapted	— .46
48	T 3 Rear. Animals	— .56

[2] This method of statistical treatment was suggested by Prof. T. L. Kelley, now at Leland Stanford University.

and the statistical method used to obtain the coefficients are all indicated.

Some of these coefficients are markedly high, also they are, almost without exception, negative.

Table 12 presents coefficients which were obtained from combining in various groups the tests which singly gave the highest coefficients. (The coefficient for NVA3—2′ time limit had not been computed at the time these groupings were made and the coefficients of Table 12 calculated.)

The general tendency, it will be noticed, is to increase the correlation coefficient by such combinations of tests.

TABLE 12

GOOD AND POOR SALESCLERKS' TESTS COMBINED

Number of Cases	Test	r by U Formula
41	Sum V A 3 N VA 3 — 3′	— .66
39	" " " 2–3–5–6 N V A 3 — 3′	— .69
43	" " " 3–5	— .79
43	" " " 2 and 3	— .84
44	" " " 2 and 5	— .78
42	" " " 2 and 6	— .56
44	" " " 2 and N V A 3 — 3′	— .66
42	" " " 3 and 6	— .64
43	" " " 2–3–5	— .84
42	" " " 5 and 6	— .51
44	" " " 5 and N V A 3 — 3′	— .61
39	" " " 6 and N V A 3 — 3′	— .75
42	" " " 2–3–6	— .79
40	" " " 2–3 N V A 3 — 3′	— .66
42	" " " 2–5–6	— .68
40	" " " 2–5 N V A 3 — 3′	— .66

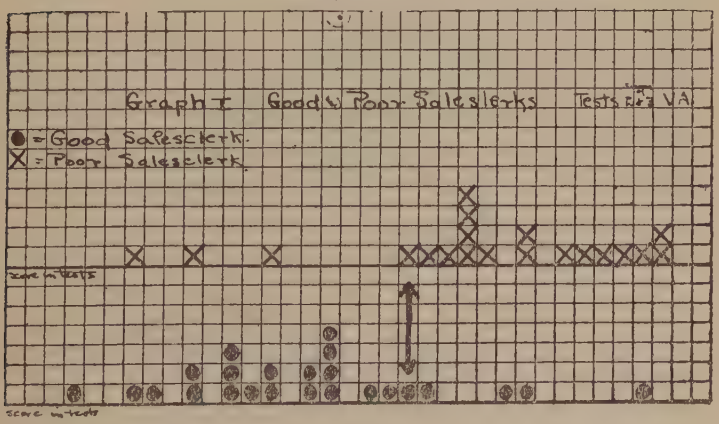


Fig. 4.

Graphically represented these coefficients take on a more concrete meaning.

Let us consider for a moment the practical implication of such a situation as Fig. 4 represents. Suppose that the forty odd people represented by the sum of the crosses and the rows of dots were, without any distinguishing mark, seated in a room. By means of tests which are represented in Fig. 4 (using the position of the arrow as a reference point) an individual would be able to determine which of the group had been successful and which unsuccessful as salesclerks, and would be in error only 8 times out of the forty-four cases.

Clerical. The clerical workers, good and poor, represented by classifications 1 and 7-9 were tested with the following tests,

V A—All
N V A—All
W. W.—Cancellation
W. W.—No. gr. checking
Rearrangement of Animals—T3

Table 13 which follows presents the coefficients which these tests gave, the formula in use being again that of unlike signed pairs.

TABLE 13

Number of Cases	Test	r by U Formula
59	N R C – V A	.34
61	" " " " " 1 Arithmetic	.09
61	" " " " " 2 Directions	.00
60	" " " " " 3 Sentences	.31
58	" " " " " 4 Synonym Antonym	.28
60	" " " " " 5 Judgment	.37
50	" " " " " 6 Analogies	.28
48	" " " N V A	.31
52	" " " " " " 1 Picture Completion	.13
52	" " " " " " 2 Series Completion	.22
51	" " " " " " 3 Comparison 2'	.45
50	" " " " " " 4 Symbol Digit	.37
52	" " " " " " 5 Form Combination	.00
34	No. Gr. Ch.	.31
35	Cancel.	— .13
57	T 3	.19

It is evident at once that the ratios are not as high as those obtained for the salesclerks, the highest being .45 for N V A 3 the Comparison Test.[3] However, it is also immediately evi-

[3] It will be remembered that this test gave the highest correlation coefficient, but a negative one, when used with salesclerks.

dent that these ratios are all positive, with the one exception of Cancellation. Now it will be remembered that the N V A and V A tests were the tests which gave the highest correlations for salesclerks but that these ratios were almost all consistently *negative.*

It seems fair to assume that had it been possible to secure clerical workers rated with the freedom from error of the salesclerks' classifications, and in view of the fact that a positive relationship is indicated by the plus ratios, that positive correlations perceptibly higher than those actually found would have been obtained. However, this is merely an assumption; what is apparent is the tendency for salesclerks and clericals to pull in distinctly opposite directions, so that we have a difference between trades as distinct as the differences within a trade, if not more so.

This is a not unimportant consideration, for the primary interest of an employment office is not in deciding with what degree of excellency an unknown individual will perform a given task, but to decide at what task to place a given individual.

Concretely, Mary Jones, age 18, without experience or training wants a job. The employment manager wants to know whether to make Mary Jones a salesclerk or a clerical. To him whether Mary Jones will rank fifth or first as a salesclerk is a matter of academic interest only. How good she will be is a matter she must show by performance, since in industry as elsewhere reward comes not in anticipation of work but follows it. Wages in a large industry are fairly standardized at employment time, and are modified not for potential ability but only for previous experience. They become really differentiated for individuals only with the history of an individual's work.

In order to express numerically the contrary tendencies of salesclerks and clerical workers, the two groups were thrown together and an attempt was made to separate from this heterogeneous group those who were called good salesclerks and good clericals.

Such a combination seemed to have two virtues, (1) the reliability of the coefficients obtained would be increased because the size of the group would be doubled; (2) the heterogeneity of the enlarged group more nearly approximated ac-

tual working conditions. Candidates for employment as they present themselves at the office which these tests were to serve are not divided off into sales or clerical workers and often have no strong preference for any special kind of work, but must be classified and selected for various jobs by the individuals who interview them.

This combination of both types of workers was made in two ways, first all sales and all clericals were combined, and selection made, for (1) good salesclerks, (2) good clerical workers. The second grouping was as follows. When good salesclerks were being selected the unsatisfactory clerical workers were omitted from the total group of sales and clericals for the reason that there did not seem to be any ground to consider them as undesirable candidates for selling.

All that was known was that they were undesirable for clerical work. Of course it may be argued that the retention of the good clerical workers was likewise ungrounded; that there was no reason why they should not make good salesclerks, even if they were good clerical workers. This is doubtless true. At the same time, considering again the practical aspect, they were actually non-sellers, the negative-positive correlation findings seemed to indicate that also potentially they might be non-sellers, and it seemed to be of interest to see what would happen to the coefficients already obtained when such a group as this under discussion was made.

When the selection for good clericals was made, the unsatisfactory salesclerks were omitted for the same reasons given for omitting unsatisfactory clerical workers when selecting for salesclerks.

Tables 14 and 15 present the coefficients which were obtained from this double grouping. The coefficients which were obtained from the pure sales and pure clerical groups are also given for comparison, so that it may be seen how the coefficients are affected by the increased number of subjects and the different kinds of groups.

Apparently, the elimination of the bad clericals, in the case of the selection for salesclerks, was justified, for when good salesclerks were selected from the heterogeneous group of all salesclerks and all clericals correlations were lower than those obtained with the salesclerks only. However when the poor clericals are omitted from this heterogeneous group and selection then made for good salesclerks the situation appears to

TABLE 14

Selecting Good Salesclerks

Test	All Sales and All Clericals		All Sales and Good Clericals Only		Good and Bad Sales Only	
	Number	r by U Formula	r by U Formula	Number	r by U Formula	Number
N R C–V A	102	—.31	—.56	71	—.59	44
" " " " " 1	104	—.22	—.28	75	—.16	43
" " " " " 2	105	—.43	—.56	76	—.66	44
" " " " " 3	103	—.59	—.75	74	—.79	43
" " " " " 4	102	—.12	—.22	73	—.22	44
" " " " " 5	105	—.48	—.64	74	—.79	45
" " " " " 6	103	—.56	—.69	74	—.54	43
" " " N V A 3–2'	95	—.96	—.93	66	—.83	44
Rear An.–T3	106	—.19	—.31	80	—.56	48
V A 3–5 N V A 3–2'	93	—.82	—.91	65	....	..
V A 3–5–6 N V A 3–2'	92	—.92	—.94	64	....	..
V A 2–3–5–6 N V A 3–2'	92	—.69	—.90	64	....	..

TABLE 15

Selecting Good Clericals

Test	All Clericals And All Sales		All Clericals and Good Sales Only		Good and Bad Clericals Only	
	Number	r by U Formula	r by U Formula	Number	r by U Formula	Number
N R C–V A	102	+.28	+.45	82	+.34	58
" " " " " 1	104	+.10	+.16	85	+.09	61
" " " " " 2	105	+.03	+.19	85	+.00	61
" " " " " 3	103	+.31	+.54	84	+.31	60
" " " " " 4	102	+.18	+.27	82	+.28	58
" " " " " 5	105	+.54	+.61	86	+.37	60
" " " " " 6	103	+.31	+.42	84	+.28	60
N R C–N V A 3–2'	95	+.66	+.64	74	+.45	51
Rear An.–T3	106	+.16	+.34	81	+.19	58
V A 3–5 N V A 3–2'	93	+.54	+.71	76	....	54
V A 3–5–6 N V A 3–2'	92	+.56	+.79	75	....	..
V A 2–3–5–6 N V A 3–2' ...	92	+.54	+.77	75	....	..

have cleared again, for the coefficients regain their original character, occasionally being slightly higher or slightly lower than in the original classification. Their significance is however increased, since their P. E.'s are decreased, because of the increased size of the group from which they were obtained.

When good clericals were selected the combining of all salesclerks and all clericals does not seem to affect the coefficients markedly, except to raise them in several instances. With the clericals, it will be remembered, the waters were originally muddy, so that an addition of several more misplacements would probably not have the same disturbing effect as with the salesclerks, where the original groupings were markedly clear.

However, when from the total group the unsatisfactory salesclerks are withdrawn and selection then made for good clerical workers, the coefficients are, with the exception of one case where the ratio remains practically the same, raised to a marked degree, and again we seem to have a justification for having excluded the poor salesclerks from this group.

Not only are the coefficients raised but in some cases they become large enough to promise a fairly satisfactory basis for the selection of clerical workers, something which the size of the coefficients obtained with the original group of clericals only did not warrant.

Final Phase of the Experiment. Having, in the manner above described, found several tests which seemed sensitive to selling and clerical ability, it became necessary, in order to employ these tests for the guidance of the Employment Office, to evaluate each test, in accordance with its importance in the group of tests. In other words, it was necessary to determine the regression equation for the whole group of tests, and the best weighting for each test in that group.

Most of the tests however which gave significant coefficients were in the group of the National Research Council Tests. These tests had been constructed for use in schools and their content was not especially adapted for industrial use.

People who are looking for a job usually object to being asked who Black Beauty or Thomas Jefferson are and are apt to ask in response "Is this a school we are in?" Even if this response is not audible, such questions do not result in a very favorable state of mind, and since the success of industrial tests is as dependent upon their reception by the people who

take the tests as upon the prophetic accuracy of the tests themselves, new tests were constructed, in form the same as certain of the N R C Tests, in substance different. The atmosphere savored more of calico, pins and sealing wax and less of George Washington, Abraham Lincoln and the states of the Atlantic Seaboard.

The test questions were all placed on one side of a sheet of 8½x11 paper, on the reverse side of which was a dotted line for the name, and several sample questions of the test itself, the samples being of course so simple as to be self-explanatory. The questions on the test paper were arranged roughly in an order of increasing difficulty, the earliest questions being scarcely more difficult than the sample questions.

A term used by Link, "shock absorber," expresses a perfectly sound principle. Non-academic groups have a very strong aversion not only to academic problems, but mental gymnastics of any kind, and it is much better for the prestige of the test to get such people started on primer tasks and then imperceptibly involved in the more difficult tasks, than immediately to frighten and make them hostile with the more difficult questions.

It was necessary, of course, after the new tests had been made to ascertain if they would act in the same way as the tests they were meant to replace, that is, give the same characteristic correlation coefficients.

This called for more testing. It will be recalled that in originally getting classifications for the salesclerks and clerical workers there were several groups, the very good, the very poor, the average, and the slightly better than average. Of these groups the extremes, good and poor of both workers had been tested in the original phase of the experiment. The clericals and salesclerks who were rated as average had not been tested.

It seemed unwise to test the original groups with the new tests, because, having spent some three hours in test taking they had had an amount of practice which made them no longer naïve, and also because it was inadvisable to take those who had already lost a considerable period of working time away from their departments for a further period.

Consequently the new tests were standardized on the group of untested workers, those who made records of average clerks,

or in some cases somewhat better, but who were neither absolutely first rate, nor third rate workers. One hundred salesclerks were tested and 43 clerical workers.

The new tests were modifications of the N R C Sentences—called in the Tables T 9, N R C Directions—T 8, N R C Comparison—T 13, N R C Judgment—T 12. In addition Completion Test T 1 A—10′ time limit—was used, the adapted form of Woodworth Wells Mixed Relations—T 5, and the Animal Rearrangement Test—T 3.

The formula which was first used to evaluate these tests was the Pearson biserial r, the clerical workers,[4] 43 in number being considered the special group and the 100 salesclerks being considered non-clerical workers.

The coefficients obtained with these tests and this group of workers is given below in Table 16.

TABLE 16

Test		Pearson Biserial r	Number
T 1A	Completion	.19	143
T 8	Directions	.63	143
T 9	Sentences	.45	143
T 12	Judgment	.55	143
T 13	Comparison	.74	143
T 3	Rearrangement of Animals	.11	143
T 5	Mixed Relations	.51	143

With the exception of T 3 and T1A the coefficients are high and the performance of the two groups, sales and clerical, is consistent with the performance of the groups previously tested, the salesclerks giving low scores, the clericals high scores.

T3 and T1A could therefore be omitted from any final group of tests. It was decided, in spite of the size of the coefficient of Test 5, which is W. W. Mixed Relations, to leave it out of consideration also, because of the difficulty always encountered in explaining the task and having it understood by those tested. With the other tests it was simply necessary for the subjects to see and do the sample problems, without other explanations from the experimenter. For the sake of economy in time and effort in the administration of

[4]It was at the time impossible to test more clerical workers because of the Christmas season rush.

the tests, when they should be used as part of the employment process, it seemed wise to discard this test, which had to be explained to the majority of subjects.

The next step was to find the regression equation for the group of tests which were retained and the best weighting for each test.

The formulae in partial correlation and tests weightings are derived from the Pearson coefficient $\frac{\Sigma xy}{n\sigma x\sigma y}$. The coefficients in Table 16 are those of the Biserial r, which closely approximates this r, but which is not identical with it.

As stated previously, the $\frac{\Sigma xy}{n\sigma x\sigma y}$ values for the new tests were found by giving the presence of sales ability the value 1, the absence of it or as was actually the case, the presence of clerical ability the value 0, the test scores forming the values of the second variable.

The regression equation was thus determined for 4 variables and the criterion, the variables being the scores in the Directions, Comparison, Judgment and Sentence Tests. The weighting of the Sentence Test however was so very inconsiderable that it was deemed inadvisable to retain it in the group and consequently a second regression equation was calculated in which there were only 3 variables and the criterion.

The following regression equation was obtained using the variables:

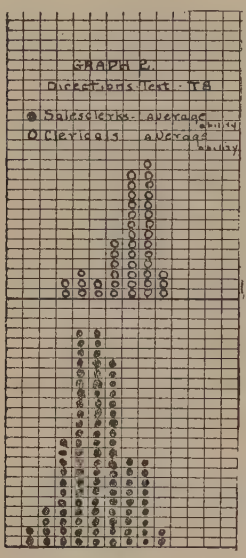
X_1 The criterion
X_2 Comparison test
X_3 Directions test
X_4 Judgment test

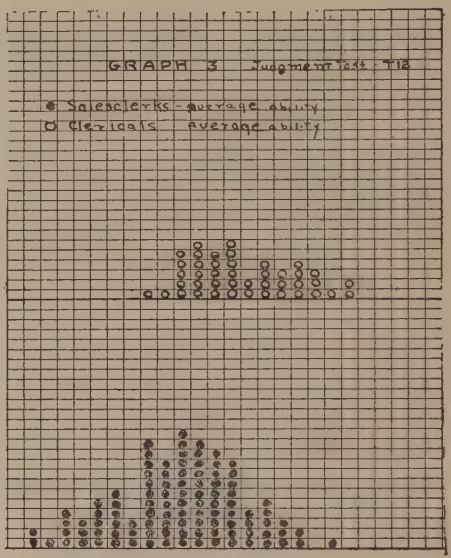
$X_1 = .025X_2 + .065X_3 - .007X_4 - .85$

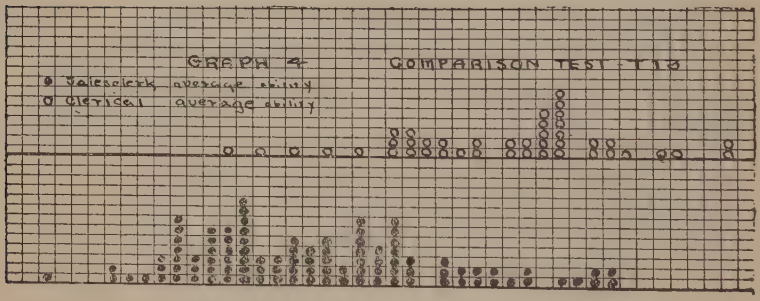
R for this equation equals .63
Biserial R equals .74

The manner in which, by weighting the tests in accordance with the regression equation above stated the division between the two groups, sales and clericals is emphasized is indicated by the following graphs.

Graphs 2, 3, 4 represent the performance of the sales and clerical groups in single tests—T18, T12, T13. The tendency for clericals to do better work is evident to some degree. A great number of misplacements would occur however if estimates of ability were made by these graphs as they stand.



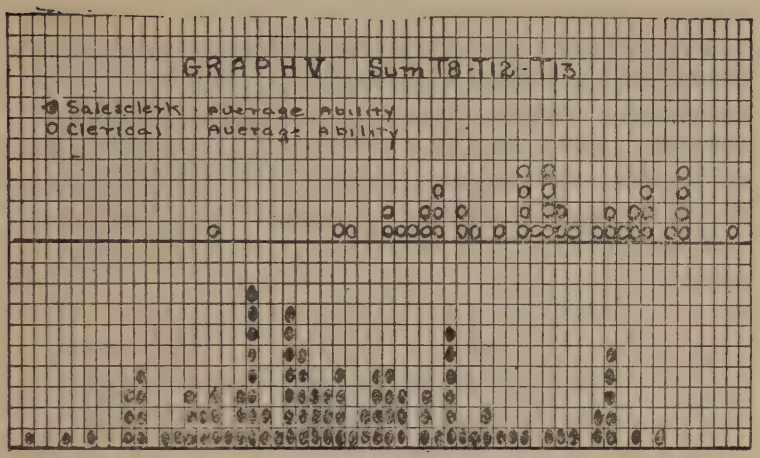




Graph V is the sum of these tests. Here the two groups show a marked tendency to separate, but there is considerable overlapping.

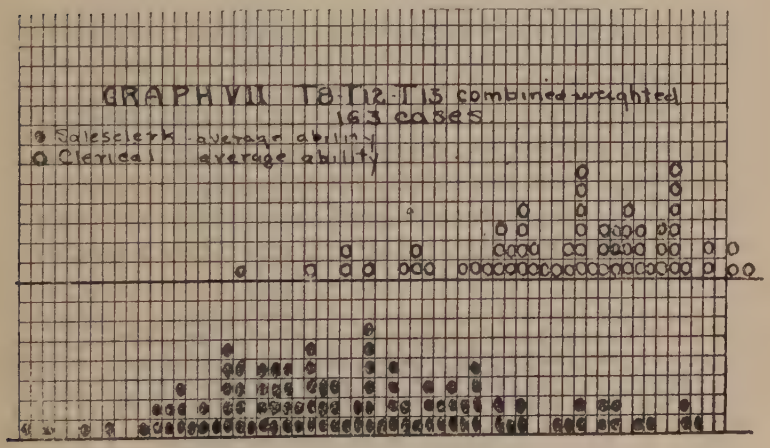
Graph VI represents the tests weighted and combined in accordance with the regression equation.

The overlapping has been diminished—and the salesclerks



form a fairly clear characteristic group, the clericals as distinctive a group.

After these regressions and other calculations had been made for the 43 clerical workers, it became possible to test 20 more clerical workers. It seemed inadvisable to recalculate on the basis of the enlarged group the weightings as originally found, but it did seem of interest to see what effect adding these 20 cases would have on the coefficient, when the tests were weighted in accordance with the values that had been obtained from the smaller group.



Pearson biserial r was raised from .74 with 143 cases to .83 with 163 cases. Graph VII shows the increased emphasis on the separation of the two groups.

Applications. All applicants for employment, during the time in which the experimental work above described was being carried on, were being tested with the Scale 1A, as described in Part I, and which served to indicate mental defectives and subnormal individuals. Upon the findings discussed above the group of tests T8, T12 and T13 were added to the test already in use and all applicants who were considered for either selling or clerical workers were tested with this group, and placement made as far as possible in accordance with the characteristic range into which the test scores fell, whether sales or clerical. The Completion Scale was still retained and in every case given to the applicant before the group of tests. Thus the fact that a low score, or one in the range characteristic of salesclerks was due to general mental inferiority, and not other causes was guarded against.

The tests have been in use for some time. In about 50% of the cases interviewers have applicants tested before placement, the other 50% of the cases are placed and then tested—the placement being subject to revision when the test findings strongly contradict the placement.

Applicants are tested generally by a clerical worker, who works under the direction of the writer and has been specially trained to give and score the tests. As many as 160 applicants have been tested in one day. Applicants are tested singly, or in groups up to 10.

The tests are also used in cases of personnel adjustment, transfers, selection of candidates for training classes, etc.

A careful follow-up by rating and production record is being carried on in order to check the test records with actual work. These are not yet in such a form that findings can be presented statistically, but on the whole, from a rough survey of the evidence at hand and from isolated cases which have come in for special attention the findings reported above tend to be borne out.[1]

[1]Subsequent to the first publication of these results the writer was able to make a rough check on the tests. Of 12 applicants employed as salesclerks whose tests indicated clerical ability, 4 were found to be completely unsuccessful, as judged by the quarterly ratings of department heads and the sales record; whereas of 58 applicants employed for the same work whose tests indicated selling ability only 6 were unsuccessful. The ratios are about 1 in three and 1 in ten.

Summary. In a large department store methods of obtaining reliable criteria for work ability of salesclerks and clerical workers were carefully developed: Workers were grouped into three classifications, very good, poor and average, and tested with a series of mental tests; only those workers being tested who were *consistently* placed in one of these three classifications by all available criteria.

The good and poor salesclerks when tested gave negative coefficients of correlation, some markedly high. Similar groups of clerical workers likewise tested gave positive correlations, not so high.

When the two groups of workers, sales and clerical, were combined, the good workers of each type could be selected from the heterogeneous group by test marks, generally with accuracy as great or greater than from the homogeneous group. This was especially true if the poor workers of one group were omitted when selecting for the good of the other.

New tests were devised which were similar in form to those that gave the highest correlations. These tests differed in content by being less academic, more industrial. These and several tests in their original form were given to a group of 100 average salesclerks and 43 average clerical workers. The relationship between the two groups which had obtained with the original tests and the original groups tested held with the new tests and the new groups.

The regression equation was determined for the three tests which, taken singly, gave the best correlations—T8, a Direction Test, T12, a Judgment Test and T13, a Comparison Test. The best weightings for these tests when combined were thus determined.

Biserial r for this group of tests and the group of 143 workers was .74. Twenty additional cases, clerical workers, raised the coefficient to .83.

These tests are now being used in connection with the employment of sales and clerical workers.

In presenting this record of the work toward the development of tests for special ability, the writer is not unaware of the fact that much of the procedure and handling of the data varies from orthodox laboratory methods. Ideal laboratory conditions, are however rarely, if ever, possible to an industrial psychologist, who must make what shift he can.

The exigencies of the occasion are therefore offered as an excuse for the departures from the orthodox. That in spite of necessarily improvised methods certain relationships were seen to hold consistently true, might in itself be offered as an "end justifying the means" excuse.

The findings are presented, for comment and criticism, and with the thought that, in the light of the present scarcity of work on tests for special abilities, the greater scarcity of literature on the same subject, and a certain feeling of discouragement which has recently become associated with work in this field, the history of this work would possess some interest.

NOTE.—It should be noted in connection with the individuals tested in this experiment, that they possessed one attribute in addition to that of ability or disability at a special job, that attribute being *permanency* as workers. The workers were rated in June, sales records were begun from May, the testing lasted into January of the following year and ratings were unobtainable for workers who had not been employed for some length of time previous to June. Consequently only such individuals could have been included among those tested as were fairly permanent workers. Stability is as important a quality in a satisfactory employee as ability itself.

It is conceivable that any superior individual, making in the tests a score beyond the range of both clerical workers or salesclerks would do excellent work at either job. It is however (since promotion can of necessity come only to a limited number), highly inconceivable that such an individual would remain for more than the briefest period at such work as department store salesclerks or clerical workers are called upon to perform, day after day, and such individuals could hardly be counted on to make up the bulk of the 5,000 workers who carry on the business of such an organization.

CHAPTER II—THE NEGATIVE CORRELATIONS

The negative correlations which were found to exist between sales ability and test performance are sufficiently unique to merit an attempt at an explanation.

The writer believes these negative correlations result from a combination of two sets of circumstances; the first, that the body of employees in this organization are a selected group from which the more able have eliminated themselves because of working conditions, and the second, that certain character traits which the tests do not measure play an important role in successful selling of the type under consideration.

In what manner these two sets of circumstances might work together to bring about an inverse relationship between sales ability and test performance is more fully considered in the following paragraphs.

It is coming to be more and more fully recognized at the present time that the tests in common use and which measure what is known as general intelligence do not thereby measure the total personality; that there are certain components left unmeasured to which the name of character traits is commonly ascribed.*

If we call a certain combination of such character traits by the name "social aptitude" then, with respect to two components of the total personality, general intelligence and social aptitude, any population can be arranged in a fourfold classification thus; (calling general intelligence G. I. and social aptitude S. A. and using plus and minus with respect to the central tendency of each component).

1– Those with plus G. I. and plus S. A.
2– Those with minus G. I. and plus S. A.
3– Those with plus G. I. and minus S. A.
4– Those with minus G. I. and minus S. A.

Even if the two qualities are considered to be positively

*R. S. Woodworth, Dynamic Psych. Chapter 8.
E. L. Thorndike, Intelligence and Its Uses, Harper's Mag., 1920, CXL, 227-235.
A. T. Poffenberger, Measures of Intelligence & Character, J. of Phil, 1922, XIX, No. 10, 261-266.
Symposium on Intelligence J. Ed. Psych. 1921, Nos. 3, 4, 5.
1921, Nos. 3, 4, 5.

correlated to the extent of .50 it may be computed from a table of the distribution of the arrays of successive tenths of the group, when r equals .50 that of every 1,000 individuals there will be 337 who will have unlike signs in the two traits.*

Now let us make the assumption that sales ability is positively related to that combination of character traits to which we have chosen to give the name of social aptitude.

Recent work upon salesmen, although not upon department store salesclerks, lends some color to this hypothesis. **M. J. Ream found that a series of the Downey will-temperament tests made over into a group test are of positive value in predicting success in selling insurance.

***B. V. Moore, working with graduate engineers, reports that in his avocational interests " . . . . the sales type of man was more interested in the social sports." The same writer further found that at college the sales engineers did their best work in social sciences.

The present investigation shows that sales ability is negatively related to tests that have the general character of intelligence tests. Then referring to the fourfold classification made above with respect to the two attributes, intelligence and social aptitude, a good salesclerk, in accordance with the assumption that social aptitude is positively associated with sales ability, can be represented by the symbols –G. I. + S. A.†

In speaking of a "good" salesclerk, particularly when the criterion for goodness is based upon ratings, one fact should not be neglected. A "good" salesclerk, from the point of view of the management, is a salesclerk who not only is able to sell, but also is *content* to sell. The satisfied worker is the stable worker. Kelley has called attention to this factor of satisfyingness when he says that†† the successful pursuit of a trade

*Computed from a set of unpublished tables received from Professor E. L. Thorndike, showing the distribution of the arrays of successive tenths of the group for various values of r.

**M. J. Ream, Group Will Temperament Tests, J. Ed. Psych, 1922, XIII, 7-16.

***B. V. Moore, Personnel Selection of Graduate Engineers, Psych. Monog. Vol. XXX, Whole No. 133, 1921.

†In connection with locating the salesclerks in the negative quadrant with respect to intelligence it might be noted that Moore, in his study (see above) found that of all the engineers tested by him with a general intelligence test, the lowest scores were made by the sales engineers.

††T. L. Kelley, Principles Underlying the Classification of Men. J. Applied Psych., 1919, XIII, 50-67.

demands among other things a certain willingness to perform the duties of the trade. And this willingness can be considered to extend, not only to the duties themselves, but also to the conditions under which the duties must be performed. If the conditions are disagreeable or rigorous in any way, if wages are low or the working day and week long—and these conditions at least may be considered as existing in the present instance—then it seems reasonable to conclude that of any group of individuals employed to work under such conditions, the most able will be the least satisfied. By reason of their superior ability they will tend to be the most able to secure jobs elsewhere either more agreeable in themselves or under more favorable conditions.

The scores on Scale 1A-10 minute of several groups of salesclerks classified according to their length of employment show this actually to be the case. Table 17 gives the median and range of the middle fifty per cent for salesclerks who had resigned after only 1, 2 or 3 months of employment and for a group of salesclerks who were still working after an interval of two years or more. The difference between the scores of the stable and unstable groups is very noticeable.

TABLE 17

Showing Duration of Employment and Scale 1A-10′ Scores

	Q1	M	Q3	No. of cases
1 month	33	37	41	119
2 months	32	39	43	34
3 months	30	35	40	60
2 yrs. or more	25	29	35	41

Now returning to the fourfold classification previously made on the basis of the attributes, general intelligence and social aptitude, and to the hypothesis that selling ability is positively correlated with social aptitude, consider the first classification, the plus G. I. plus S. A., made up of individuals who are positive with respect to the C. T. in both qualities. Such individuals represent the most able of this fourfold classification, for they, and they alone possess both qualities in a positive degree.

Following the data and argument concerning the relative stability of the less and more able employees, such individuals

will tend to eliminate themselves most rapidly from this organizations. They will do so not because they are not able to perform the duties of the trade, but because they are the least willing to perform these duties under the conditions which obtain in the trade, or because the duties themselves are irksome.

The group classified as –G. I. –S. A. on the other hand, who are unfortunate enough to possess neither intelligence nor the necessary character traits, will not tend to eliminate themselves, perhaps, but will be the most rapidly eliminated by the management.

Thus the two groups which are left as the most stable are the groups with unlike signs with respect to the two attributes under consideration. It has already been shown how one of these, the –G. I. +S. A. could be considered to represent the good salesclerks. For the same reasons the other group, the *+G. I. –S. A. can be considered to represent the poor salesclerks. If two such groups were tested with tests which measured only the factors entering into general intelligence, it is plain that a negative relationship would appear; the good salesclerks would do poor test work, the poor salesclerks good test work.

That the subjects of this investigation were essentially stable employees, the footnote on page 48 points out. That they had, therefore, a certain contentment with the situation they were in would seem naturally to follow. There is direct evidence—see Table 17—that individuals of superior ability, at least so far as performance on Scale 1A indicates it, tend to be less stable in this organization and at this job.

It is the writer's belief that the subjects of this investigation actually did belong to two groups which were of unlike sign with respect to the attributes general intelligence and social aptitude, and that the negative correlations resulted from the fact that these groups were tested with tests which measured only the first of these attributes. It remains for further investigation to determine if another kind of intelligence, that related to social aptitude, occurs in these or similar subjects in accordance with the hypothesis.

*It might be expected that of these two groups the +G. I. –S. A. would be the less stable. Such, in fact, was the case, for during the course of the investigation there was great difficulty in reaching the salesclerks who had been classified as poor in time to test them. Although employed long enough to receive ratings and a production record, they were constantly leaving in the interval that elapsed before testing.

NOTE: The fact that working conditions exercise a certain influence on the character of the group of workers has certain bearings upon the results that may be expected from investigations within the same organization at different times or within different organizations.

If the groups are of the same character, similar results, of course, may be expected. But any change in the management or in policy, an increase or decrease in wages or in working hours is likely to be reflected in the selection of workers. So trivial a matter as a generous or parsimonious use of fresh paint would not be without its effect.

The use of tests themselves for selective purposes is no small factor in changing the character of the workers. That is indeed their purpose. The group becomes more homogeneous with respect to whatever is measured by the test. The employment manager and his assistants are not unaffected by the use of tests even when they are not guided by them, for, when applicants are employed contrary to test indications, they are likely to be more carefully selected for that reason.

And, in so far as the labor supply of an organization is conditioned by the character of the people who already work there, that is, in so far as applicants are friends or relatives, or friends' friends or relatives' friends of employees, in so far would the use of tests change even the character of the labor supply.

It follows from these considerations that the results of an industrial research may need to be constantly revised in order to keep pace with the changing character of the personnel. The investigator will need to be cognizant of the changes as they take place so as to keep his investigation abreast of them.

CHAPTER III—THE PSYCHOLOGIST IN INDUSTRY

In the preceding chapters very little has been said about the place of the psychologist in industry. Perhaps little needs to be said. Provided his tools be adequate the psychologist can function in any situation in which a judgment concerning a human being is called for. The task is now to create and perfect tools.

The tests which have been discussed in these pages functioned chiefly in connection with the activities of employment and education. Under the former they were used in the selection of applicants for employment and their proper placement, in cases of promotion, transfer and wage adjustment; under the latter in the selection of employees for special training and opportunity classes. There were in addition a number of miscellaneous ways in which the tests were of use.

In exact form the manner in which tests will function will vary, of course, from one organization to another, depending upon such factors as its size and structure and other individual characteristics.

Tests for general intelligence have a certain usefulness in the industrial situation. They may be used to set limits between which would lie the optimum degree of intelligence for any given occupation, when all the factors which combine to make a satisfactory employe are taken into consideration. In any case, a knowledge of the general level of ability of an individual seems to be fundamental to intelligent action concerning the individual. A priori generalizations, however, upon the course of action based upon such knowledge should of course not be made. For witness the fact that in these studies, in different situations, postive and negative association, and the absence of association, have all been found.

In choosing or constructing tests for industrial use the investigator will generally need to be mindful of certain considerations in addition to customary ones. These involve chiefly the time factor, the simplicity of administration and scoring ,the relevancy of the test content to the industrial situation, and the difficulty of the task as related to the intellectual status of the individuals tested.

In evaluating the tests it will probably be necessary to be at some pains to make the criteria of ability reliable. Very few organizations have personnel records that are acceptable to the psychologist. Yet a true evaluation of the tests is directly conditioned by the reliability of the criteria. The first work of the psychologist is then very likely to involve the creation of a system of personnel and production records, or the radical revision of an existing system.

Such a system of records, evolved as a by-product to the evaluation of certain tests and for the use of the psychologist, is also of interest and value to the organization in general. In developing them the psychologist has a field of usefulness that is in addition to the application of tests.

APPENDIX

SCALES 1 AND 1A

Scale 1

1. We like good boys ..... girls.
2. Men ..... older than boys.
3. I like to go to .....
4. The kind lady ..... the poor man a dollar.
5. Good boys ..... kind .... to their sisters.
6. Boys and ..... soon become ..... and women.
7. Time ..... often more valuable ..... money.
8. The poor baby ..... as if it were ..... sick.
9. Children should ..... many lessons from ..... parents.
10. The child ... ... the river ..... was drowned.
11. The ..... are often more contented ..... the rich.
12. She ..... if she will.
13. It is good to hear ..... voice .... friend.
14. ..... the weather is ..... one should wear heavier ..... than when it is .....
15. The poor little ..... has ..... nothing to .....; he is hungry.
16. Worry ..... never improved a situation but has ..... made conditions .....
17. It is very ..... to become ..... acquainted ..... persons who ..... timid.
18. The ..... of your ..... and mother is your brother.
19. One's ..... do ..... always express his thoughts.
20. To ..... many things ..... ever finishing any of them .... a ..... habit.
21. The knowledge of .... use fire is ..... of ..... important things known by ..... but unknown ..... animals
22. The ..... seems ..... and dreary .... conditions .....
23. ..... that are ..... to one by an ..... friend should be pardoned ..... readily than injuries done by one ..... is not angry.
24. In order .... clearly at ..... it is ..... to ..... artificial .....
25. It is ..... that a full-grown man .....see a ghost ..... he is .....

Scale 1A

1. The sky ..... blue.
2. Ice is cold, but fire is .....
3. I ..... to school each day.
4. The ..... plays ..... her dolls all day.
5. The girl fell and ..... her head.
6. The wind ..... the dust into our eyes.
7. The boy will ..... his hand if ..... plays with fire.
8. The ..... rises .....the morning and ..... at night.
9. The boy who ..... hard .....do well.
10. Hot weather comes in the ..... and ..... weather ..... the winter.
11. Children ..... to pick .....
12. A ..... drink is very refreshing to a ..... who is .....

13. It is a .....task to be kind to every beggar .... for money.
14. Men ..... more ..... to do heavy work ..... women.
15. It is hard .....keep ..... getting ..... on a rainy day.
16. When one ..... angry he should ..... forth an effort .... his actions.
17. In ..... to maintain ..... health one should have nourishing .....
18. A home is ..... merely a place ..... one ..... live comfortably.
19. Sleep ..... both ..... and body.
20. The ..... is always shining ..... storm clouds sometimes ..... it ..... us.
21. When two persons ..... about ..... which neither understands, they ..... almost ..... to disagree.
22. Extremely old ..... sometimes ..... almost as ..... care as .....
23. It is sometimes ..... to ..... between two ..... of action.
24. The least difficult ..... are by no ..... always the most ....., ... are the ..... tasks ..... the most disagreeable.
25. ..... they ..... us ..... not, nature's ..... are ..... and unchangeable.

BIBLIOGRAPHY

Bills, M. A. Methods for the Selection of Comptometer Operators and Stenographers. J. Appl. Psych., 1921, V, 275-83. A Test for Use in Selection of Stenographers. J. Appl. Psych., 1921, V, 373-77.

Burtt, H. E. Employment Psychology in the Rubber Industry. J. Appl. Psych., 1920, IV, 1-17.

Flanders, Jesse Knowlton, Mental Tests of a Group of Employed Men Showing Correlation with Estimates Furnished by Employer. J. Appl. Psych., 1918, II, 197-218.

Freyd, Max. A Test Series for Journalistic Aptitude. J. Appl. Psych., 1921, V, 46-57.

Henmon, V. A. C. Air Service Tests of Aptitude for Flying, J. Appl. Psych., 1919, III, 103-109.

Hollingworth, H. L. Vocational Psychology, D. Appleton & Co., New York, 1916.

Kelley, T. L. Principles Underlying the Classification of Men. J. Appl. Psych., 1919, III, 50-67.
Individual Testing with Completion Test Exercises, Teachers College Record, 1917, XXIII, 371-383.

Kitson, H. D. Vocational Guidance and Theory of Probability, School Review, 1920, II, 143-49.

Kornheiser, A. W., Meine F., Ruml, B. Two Models Showing the Interrelation of several Significant Correlation Variables. Psych. Bulletin, 1919, XVI, 230-35.

Link, Henry C., Employment Psychology, MacMillan Company, New York, 1919.

Marcus, L. Vocational Selection for Specialized Tasks, J. Appl. Psych., 1920, IV, 186-201.

Moore, B. V. General Intelligence Determined by Its Weakest Essential Element. J. Appl. Psych., 1920, IV, 155-161. Personnel Selection of Graduate Engineers, Psych. Monograph, 1921, XXX, Whole No. 133.

National Academy of Sciences. Memoirs, 1921, XV.

National Association of Corporation Schools (Training) Psychological Tests and the Results Obtained Therefrom, Seventh Annual Proceedings, 1919. Application of Psychological Tests and Rating Scales in Industry. Eighth Annual Proceedings, 1920. Application of Psychological Tests and Rating Scales in Industry. Ninth Annual Proceedings, 1921.

Oschrin, E. Vocational Tests for Retail Saleswomen, J. Appl. Psych., 1918, II, 148-155.

Otis, Arthur S. Selection of Mill Workers by Mental Tests, J. Appl. Psych., 1920, IV, 339-341.
Otis Group Intelligence Scale, World Book Co., 1921.

Poffenberger, A. T. Measures of Intelligence and Character. J. of Phil., 1922, XIX, No. 10, 261-266.

Pearson, Karl. A New Method of Determining Correlation Between a Measured Character A and a Character B—Biometrika 1909, VII, 96-105.

Pressey, S. L. An Attempt to Measure the Comparative Importance of General Intelligence and Certain Character Traits in Contributing to School Success. Elem. Sch. J., 1920, XXI, 220-227.

Ream, M. J. Group Will Temperament Tests. J. Ed. Psych., 1922, XII, 7-16.

Rugg, H. Is the Rating of Human Character Practicable? J. Ed. Psych., 1921, XII, 425-438, 485-501, 1922, XIII, 30-42, 80-93.

Rogers, H. W. Psychological Tests for Stenographers and Typists, J. Appl. Psych., 1917, I, 268-274.
Soper, H. E. On the P. E. of the Bi-serial Expression for the Correlation Coefficient. Biometrika, 1914, X, 384-390.
Stenquist, J. L. Case for the Low I. Q. J. Ed. Research, 1921, IV.
Thorndike, E. L. Mental and Social Measurements. Teachers College, Columbia University, 1916.
Fundamental Theorems in Judging Men. J. Appl. Psych., 1918, II, 67-76.
Tests of Intelligence-Reliability, Significance, Susceptibility to Special Training and Adaptability to the General Nature of the Task, Sch. & Soc., 1919, IX, 189-195.
The Reliability and Significance of Tests of Intelligence, J. Ed. Psych., 1920, XI, 284-287.
A Constant Error in Psychological Ratings. J. Appl. Psych., 1920, IV, 1-17.
Intelligence and Its Uses. Harper, 1921, CXL, 227-235.
Thurston, L. L. Mental Tests for College Freshmen, J. Ed. Psych., 1919, X, 129-142.
Mental Tests for Prospective Telegraphers, J. Appl. Psych. 1919, III, 110-117.
Trabue, M. R. Completion Test Language Scales, Col. Univ. Contributions to Education, 177, 1916.
Woodworth, R. S. Dynamic Psychology, Col. Univ. Press, N. Y., 1918.
Yule, G. U. An Introduction to the Theory of Statistics, Charles Griffin & Co., London, 1919.

VITA

ELSIE OSCHRIN BREGMAN. Born in Newark, New Jersey, November 30, 1896. Educated in the public schools of Jersey City, New Jersey. Diploma of graduation from Dickinson High School, 1914. Barnard College, Columbia University, B. A., 1918. Statistical Division of the Quartermasters Corps, U. S. A., 1918. Psychologist, Guardian Life Insurance Company, 1918-1919. Psychologist, R. H. Macy & Company, 1919-1921. Graduate Student, Columbia Univ., Department of Psychology, 1921-1922.

Previous Publications. Vocational Tests for Retail Saleswomen, J. Appl. Psych. June, 1918.

The Application of Psychological Tests and Rating Scales in Industry. National Association of Corporation Training, Ninth Annual Proceedings, 1921.

A Study in Industrial Psychology—Tests for Special Abilities. J. Appl. Psych. June, 1921.

A Scientific Plan for Sizing Up Employees. System. June, 1922.